Supplementary Information

Testing for ancient admixture between closely related species

Eric Y. Durand[1,*], Nick Patterson[2], David Reich[2] and Montgomery Slatkin[1]

[1]Department of Integrative Biology, University of California, Berkeley, CA 94720;

[2]Broad Institute of MIT and Harvard, Cambridge, MA 02142.

[*]To whom correspondence should be addressed; e-mail: eric.durand@berkeley.edu

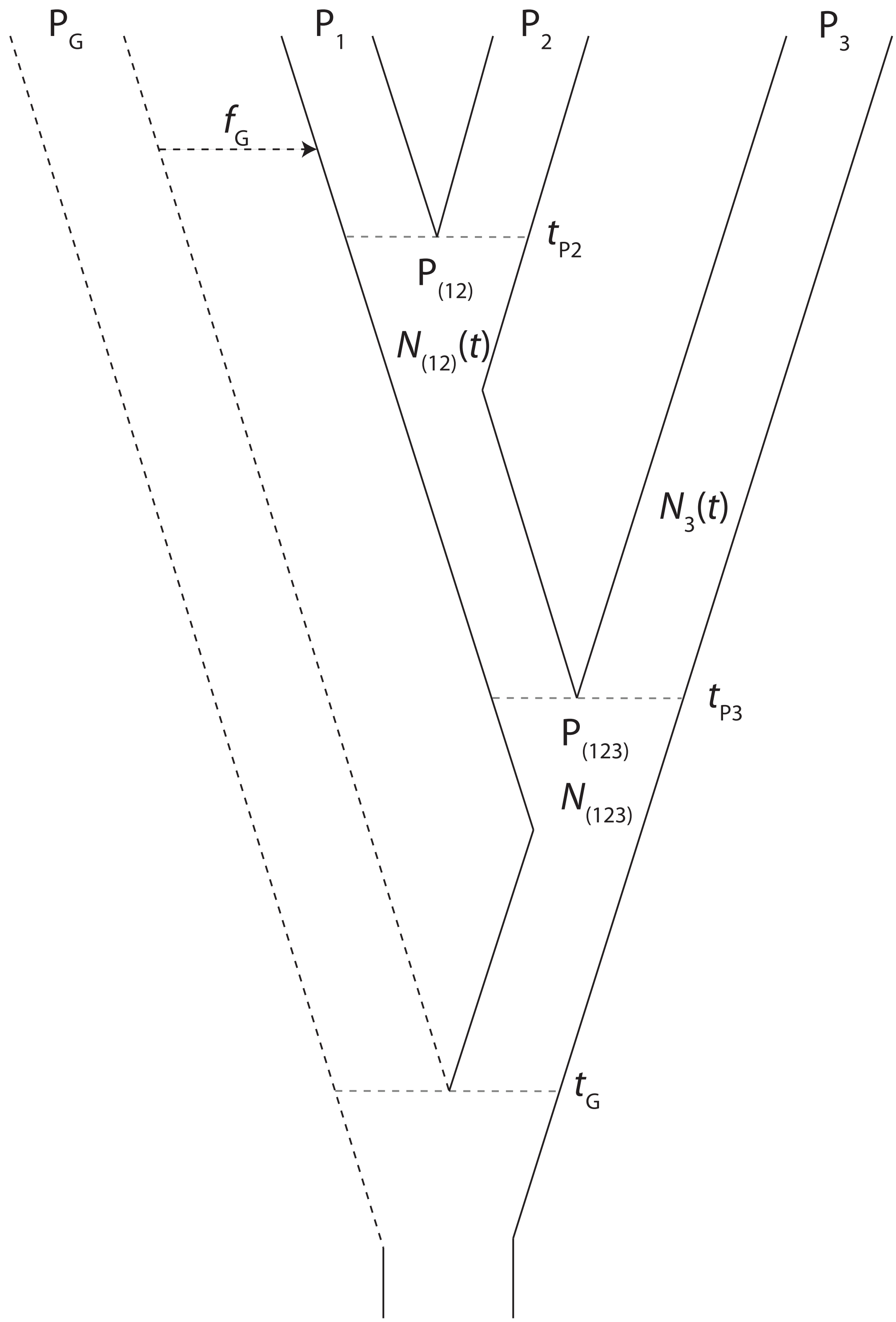

Figure S1: Model of instantaneous admixture with a ghost population. P_1 and P_2 split at time t_{P2}. P_3 split from the ancestral population of P_1 and P_2, denoted $P_{(12)}$, at time t_{P3}. An unsampled, ghost population denoted P_G split from the ancestral population of P_1, P_2 and P_3 at time t_G. A single episode of admixture takes place from P_G to P_1 at rate f_G. This model can fit the same *D*-statistics as the IUA model because, in the case where P_1 originated from P_G (probability f_G), P2 and P3 have an extra opportunity to coalesce between t_{P3} and t_G.

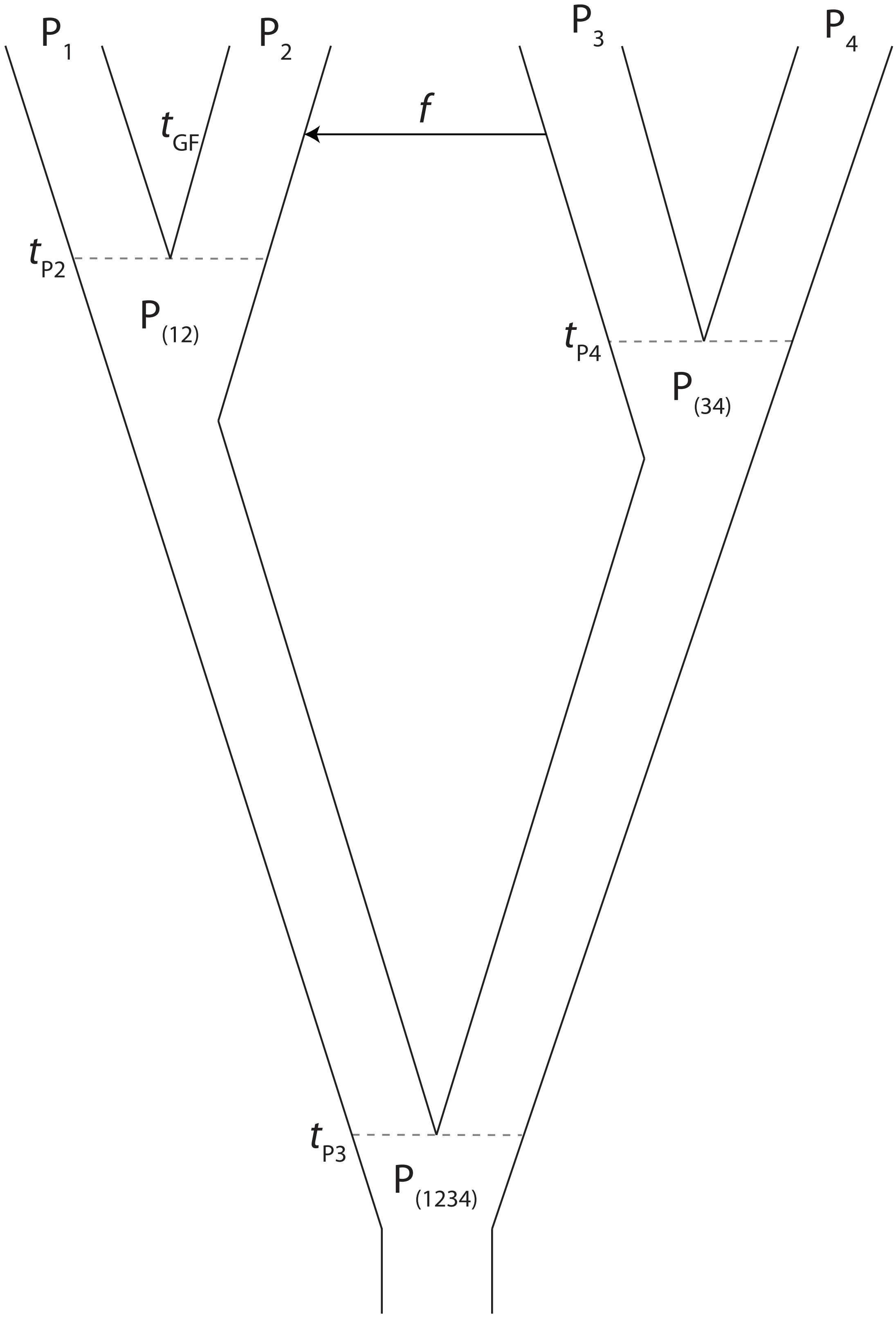

Figure S2: Model of instantaneous admixture where a sister group of P_3 is available. The sister population, P_4, split from P_3 at time t_{P4}. P_3 split from the ancestral population of P_1 and P_2, denoted $P_{(12)}$, at time t_{P3}. A single episode of admixture takes place from P_3 to P_2 at time t_{GF}. We denote the admixture proportion by *f*. The sample from P_4 leads to an unbiased estimator of *f* that does not depend on any other demographic parameters.

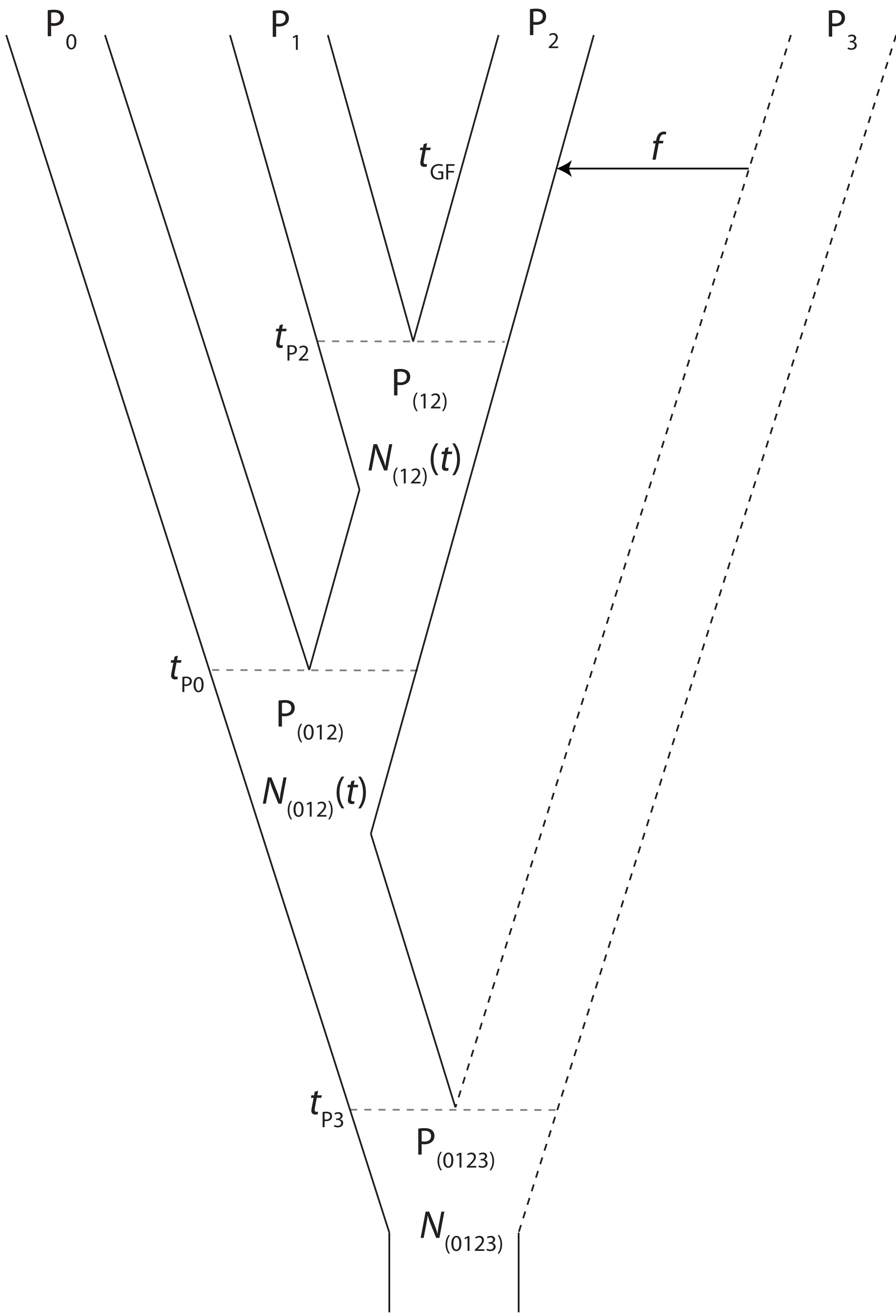

Figure S3: Model of instantaneous admixture where the admixing population has not been sampled. Instead, an outgroup to P_1 and P_2, denoted P_0, has been sampled. P_0 split from the ancestral population of P_1 and P_2 at time t_{P0}. A single episode of admixture takes place from P_3 to P_2 at time t_{GF}, but P_3 has not been sampled. The admixture proportion is f. This model allows us to estimate f without using the sample from the archaic population. In the case, because P_2 originated from P_3 (probability f), P_1 and P_0 have an extra opportunity to coalesce between t_{P0} and t_{P3}.

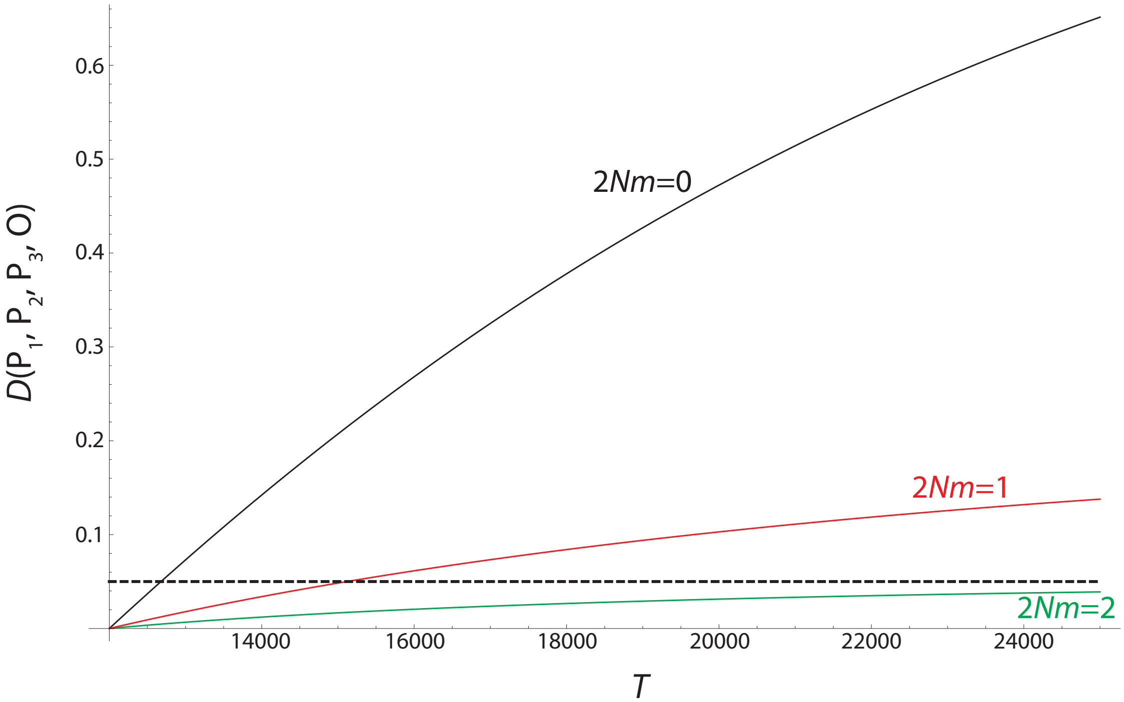

Figure S4: ***D*****(P_1, P_2, P_3, O) as a function of *T* under the AS model for different migration rates *m*, *N*=10,000, t_{P2}=3,000 generations and t_{P3}=12,000 generations. The dashed horizontal line represents *D*(P_1, P_2, P_3, O)=0.05, corresponding roughly to the observed value in (Green et al. 2010). In a model with no migration (black curve), *T* needs to be roughly 13,000 generation to match *D*(P_1, P_2, P_3, O)=0.05. The value of *T* corresponding to *D*(P_1, P_2, P_3, O)=0.05 increases quickly with *m*. Note that, under the AS model, low migration rates predict that P_1 and P_2 are more diverged than P_2 and P_3.**